楽曲特性をデータとした楽曲の印象空間の推定
ーカラオケにおける楽曲リコメンドサービスへの応用をねらいとしてー
○磯貝 悟志（慶應義塾大学） 中西 美和（慶應義塾大学）

Modeling of the music impression map using lyrics and acoustic features
-Aiming at applying to the music recommendation of the Japanese Karaoke service-
Satoshi ISOGAI (Keio University), Miwa NAKANISHI (Keio University)

1. はじめに

近年、商品やサービスの特性とユーザの特性を関連づけて、リコメンデーションに活用する方法の探索が進められている。本研究では、カラオケを対象サービスとして、ユーザの歌唱意欲を促進する楽曲の方法について検討する。具体的には、ユーザが楽曲の印象をどのように捉えるのかを明らかにし、その上で、ユーザが好む楽曲から一定の類似性を持つ楽曲をリコメンデーションする方法の確立を目指す。本稿では、その第一ステップとして、ユーザが抱く印象の観点で楽曲を位置付けする、楽曲マップの構築を試みる。ただし、実システムへの応用を鑑みたとき、個々の楽曲の印象について、ユーザの印象を問うことは現実的ではない。そこで、所与として得られる楽曲データから、ユーザの印象に基づく楽曲マップを推定するモデルの構築をする。

2. 実験

実験は2期に分けて行った。実験Ⅰでは様々なジャンルから合計80曲、実験Ⅱでは実験Ⅰとは異なる51曲を視聴対象楽曲とした。被験者には、Bメロからサビまでを約50秒に編集した楽曲を3回試聴させ、表1、2に示す質問項目に答えさせた。実験Ⅰと実験Ⅱそれぞれ異なる平均年齢22歳の学生被験者20名（内女性10名）に実験開始前に実験についての説明を行い、書面での同意を得てから参加させた。

表1 実験に用いた印象を問う質問項目

形容詞項目(0から6のフリースケールで回答)				
シンプルな	異質感	若々しい	安らぎ感	軽快感
躍動的な	頑固な	人工的	信頼感	成熟した
斬新な	力強い	レトロな	高級感	非日常的
発展的な	大衆的な	開放感	優雅な	明るい
派手な	とがった			

表情語(木通, 2004)をもとに改編

表2 実験に用いた認知度を問う質問項目

項目	回答方法	選択肢
知っている度	単一選択	1全く知らない
		2サビのみ知っている
		3アーティスト名か曲名を知っている
		4アーティスト名、曲名の両方を知っている
		5アーティスト名、曲名も分かり、曲全体もだいたい分かる

3. 楽曲の印象空間の同定

印象空間の構築にはいくつかの手法が考えられるが、本研究では、各形容詞項目に対する平均評価得点を主成分分析するという方法を採用した。累積寄与率70%以上を目安に、第二主成分までを採用すると、図1に示す主成分負荷量が、両実験について得られた。両実験から得られた主成分負荷量の傾向は非常に類似していることがわかる。実験Ⅰでは41.6%、実験Ⅱでは56.5%の寄与率となった第一主成分では、「躍動的」等の負荷量が大きく正の値を示しており、「静かさ」等の負荷量が大きく負の値を示している。従って、この主成分を「エネルギーシユさ」の成分と解釈した。一方、実験Ⅰでは28.1%、実験Ⅱでは19.6%の寄与率となった第二主成分では「大衆的」等が大きく正の値を示しており、「異質」等が大きく負の値を示している。従って、この成分を「親しみやすさ」の成分と解釈した。全く異なる楽曲、被験者を用いた二つの実験において、同じ主成分を見いだせたことから、この二つの成分を軸として構成される楽曲の印象空間は、楽曲やユーザの種類によらず普遍性を持つものと見なすことができる。

図1 実験Ⅰ、実験Ⅱ主成分負荷量

4. 楽曲特性に基づく楽曲マップの推定

冒頭に述べたように、無数の楽曲について、ユーザが試聴し、楽曲マップを作ることは、非現実的である。そこで、任意の楽曲を、それらに付随する所与のデータによって、印象空間上に位置付

けるためのモデルを構築する。なお、以下では、実験Ⅱで得られたデータを用いてモデルを構築し、実験Ⅰで得られたデータを用いてモデルを検証する。

4.1 「エネルギー感」の量的推定

実験Ⅱのデータについて、前章で得られた51楽曲の第一主成分「エネルギー感」に対する主成分得点（以下「エネルギー感得点」とする）を目的変数とし、それを楽曲の音響特徴量で推定するモデルの構築を試みた。まず、代表的な5つの音響特徴量（表3）をコードブックサイズ16でベクトル量子化して80の変数を生成し、そのうち「エネルギー感得点」と相関の高い6変数を説明変数として採用した。最小二乗法によるパラメータ推定の結果、得られたモデルを式（1）に示す。

$$E = -0.43 \times \sqrt{x_1} - 0.19 \times x_2 - 0.37 \times \log(x_3) + 1.08 \times \log(x_4) - 0.52 \times \sqrt{x_5} + 1.60 \times \log(x_6) + 2.65 \dots (1)$$

E:エネルギー感得点

x_1 :MFCCをベクトル量子化した際の代表ベクトル14の割合

x_2 :セントロイドをベクトル量子化した際の代表ベクトル3の割合

x_3 :ブライトネスをベクトル量子化した際の代表ベクトル12の割合

x_4 :ブライトネスをベクトル量子化した際の代表ベクトル13の割合

x_5 :ロールオフをベクトル量子化した際の代表ベクトル3の割合

x_6 :クロマベクトルをベクトル量子化した際の代表ベクトル3の割合

表3 5つの音響特徴量

音響特徴量	概要
MFCC ²⁾	スペクトル包絡を表現する特徴量。本研究では13次元までを使用する。
スペクトルセントロイド	各時刻におけるスペクトルの重心。
ブライトネス ³⁾	高周波数エネルギーとも呼ばれ、特定の周波数よりも高い帯域に全体の何%のエネルギーが含まれるかを計算する。本研究では1500Hzを基準とした。
スペクトルロールオフ	低周波数側から振幅スペクトルの累積和をとった際に累積和の値が85%に達する周波数である。
クロマベクトル ⁴⁾	12音名の各音名(ピッチクラス)の周波数パワーを複数のオクターブに渡って加算した12次元の特徴量

実験Ⅱ、実験Ⅰで用いた全楽曲の音響特徴量を分析し、(1)式の精度を検証したところ、実験Ⅱの楽曲では決定係数0.82、実験Ⅰの楽曲では決定係数0.66となった。このことから、(1)式は、楽曲に対してユーザが感じる「エネルギー感」を各楽曲の音響的な特徴を用いて推定するモデルとして高い精度を持つものと言える。

4.2 「親しみやすさ」の量的推定

実験Ⅱのデータについて、前章で得られた51楽曲の第二主成分「親しみやすさ」に対する主成分得点（以下「親しみやすさ得点」とする）を目的変数とし、それを楽曲の認知度及び歌詞特性で推定するモデルの構築を試みた。楽曲の認知度は、70%以上の被験者が一部分でも知っているとした場合1、そうでない楽曲を0として数量化した。また、歌詞特性には、音楽における表現に適した感情表現語を8つの群に分類した「Hevnerの形容詞環」

を用いた。各楽曲の歌詞を単語単位に分割し、含まれる形容詞8群のいずれかに分類して、各群について含有率を求めた。表4にある楽曲の各群の含有率を例として示す。

表4 楽曲中に含まれる形容詞の8群への分類結果（上段は数、下段は割合）

	関連語群1	関連語群2	関連語群3	関連語群4	関連語群5	関連語群6	関連語群7	関連語群8	合計
楽曲1	0	1	1	0	0	4	1	3	10
楽曲1 (割合)	0	0.1	0.1	0	0	0.4	0.1	0.3	1

最小二乗法によるパラメータ推定の結果、得られたモデルを式（2）に示す。

$$F = -2.23 \times x_1 - 1.95 \times x_2 + 0.04 \times x_3 + 0.37 \times x_4 + 0.23 \times x_5 + 0.23 \times x_6 - 0.70 \times x_7 + 0.84 \times x_8 - 0.07 \dots (2)$$

F:親しみやすさ得点

x_i :各群の形容詞 j の含有率($i=1\sim7$)

x_8 :認知度(1, 0)

実験Ⅱで用いた全楽曲及び実験Ⅰで用いた楽曲のうち外国語曲などを除く40曲の歌詞を分析し、(2)式の精度を検証したところ、実験Ⅰの楽曲では決定係数0.62、実験Ⅱの楽曲では決定係数0.53となった。このことから、(2)式は、楽曲に対してユーザが感じる「親しみやすさ」の印象を各楽曲の歌詞及び認知度を用いて推定するモデルとして、比較的高い精度を持つものと言える。

5. まとめ

本研究では、ユーザが好む楽曲から一定の類似性を持つ楽曲をリコメンデーションする方法の確立を目指して、楽曲の印象を楽曲の持つ特徴量から推定するモデルを構築した。

参考文献

- 1)Hevner.K:“Experimental studies of the elements of expression in music”, The American Journal of Psychology, Vol. 48, No. 2, pp. 246-268, 1936.
- 2)西村義隆:“周波数帯域ごとの重みつき尤度を用いた雑音に頑健な音声認識”, 音声言語情報処理 2003(124), pp.19-24, 2003.
- 3)P.N.Juslin:“Cue utilization in communication of emotion in music performance: relating performance to perception”, Journal of Experimental Psychology: Human Perception and Performance, Vol. 26, No. 6, pp. 1707-1813, 2000.
- 4)Mark A Bartsch et al:“To catch a chorus: Usingchroma-based Representations for Audio Thumbnailing”, WASPAA'01, pp. 15-18, 2001.